

CERTIFICATE OF ELECTRONIC TRANSMISSION UNDER 37 C.F.R. §1.8

I hereby certify that this correspondence is being filed electronically with the U.S. Patent and Trademark Office on the below date:

Date: 6/19/09 Name: James P. Naughton

Signature: 

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of:

Shingo Kiuchi, et al.

Serial No. 10/730,767

Filing Date: December 8, 2003

For SPEECH RECOGNITION PERFORMANCE  
IMPROVEMENT METHOD AND SPEECH  
RECOGNITION DEVICE

Attorney Docket No. 9333/361

Examiner: James S. Wozniak

Group Art Unit No.: 2626

Confirmation No.: 3437

**REPLY BRIEF**

Commissioner for Patents

P.O. Box 1450

Alexandria, VA 22313-1450

Dear Sir:

Appellants submit this brief in reply to the Examiner's Answer dated May 14, 2009, which maintained all rejections previously appealed.

## **I. THE EXAMINER'S CLAIM INTERPRETATION IS INCORRECT**

Throughout the Examiner's Answer, the Examiner criticizes Appellants for allegedly stating that the claimed invention determines the "actual" starting point of a speech region, whereas the Examiner asserts that only a "general" start position is found. First, Appellants' description of the claimed invention did not refer to an "actual" starting point (although Appellants' discussion of the cited art did). Second, Appellants disagree with the Examiner's assertion that "only a general start position is found."

Characterizations aside, the fact is that Appellants' claims recite "identifying a start position of a speech region of speech data for which speech recognition is to be performed," period. Nothing could be plainer or clearer: a start position of the speech region is identified -- not some vague "general" start position as asserted by the Examiner. Thus, contrary to the Examiner's assertion, Appellants' invention does involve determining a starting point of a speech region of speech data, and the claims expressly recite this.

## **II. THE EXAMINER'S CHARACTERIZATION OF FUJII IS INCORRECT**

The Examiner does not dispute that a key feature of Fujii, the primary reference, is not needing to identify the starting point of a speech region, in contrast to Appellants' claimed invention. However, the Examiner wrongly asserts that Fujii teaches determining multiple candidate speech periods by adding a certain amount of a non-speech region to a most likely speech period (Examiner's Answer at pp. 7, 8). To the contrary, the cited passage in Fujii (col. 8, lines 11-49) includes the statement that: "In response to the detected power levels in the three [frequency] bands, a speech period detecting portion 43 responsive to band pass filter portion 42 extracts the periods having largest power level as proposed speech periods. In order to avoid ambiguity of the boundary due to unvoiced sounds or noise introduction, plural possible speech periods are extracted in portion 43 as the proposed periods." Nowhere does Fujii describe determining multiple proposed speech periods by adding a certain amount of a

non-speech region to a most likely speech period, as asserted by the Examiner. In fact, as conceded by the Examiner elsewhere, Fujii is silent on how specifically to determine the starting points of the multiple speech periods (Examiner's Answer at p. 9, top).

Further, because Fujii is silent on how the beginning points of the proposed speech periods are determined, there may be no non-speech region included in the proposed speech periods if the beginning points are taken after the start of a speech region. Moreover, Appellants' claims recite that, in addition to a varying period of a preceding non-speech region, the plurality of pieces of speech data all include "said speech region" of speech data to be recognized. Fujii does not describe that all of its proposed speech periods contain the speech region of speech data to be recognized. We just don't know these aspects of Fujii, because Fujii doesn't say. However, it appears from the quote above that the multiple periods in Fujii may be determined according to power levels in different frequency bands - - an approach which is very different than in Appellants' claimed invention.

### **III. THE EXAMINER'S CHARACTERIZATION OF BI IS INCORRECT**

The Examiner is incorrect on two accounts by asserting that Bi determines a "likely speech start position (PRE\_START)" from which "multiple candidate periods with different starting points" are generated (Examiner's Answer, e.g., at pp. 8, 10-12).

First, the PRE\_START position described by Bi is not a likely speech start position. As explained by Appellants (Appeal Brief at pg. 9), PRE\_START is only an interim calculation point on the way to calculating the ultimate starting point. A key feature of Bi is using two signal-to-noise ratio (SNR) thresholds to detect a start or endpoint of a valid speech region. A first, higher SNR threshold is used to capture relatively strong voice segments in the utterance and establish PRE\_START, whereas the second, lower SNR threshold is used to find relatively weak segments in the utterance (e.g., col. 2, lines 42-44; col. 4, lines 47-54). Thus, PRE\_START is not the starting point of the speech region, but rather the variable START which is obtained

after performing Bi's process "represents the actual starting point of the utterance" (col. 7, lines 24-25). This is also evident from the name "PRE\_START" which indicates that it is "pre", or before, the start, and is seen from Fig. 2 where PRE\_START clearly is not the start of the speech region.

Second, the "look back" feature of Bi does not generate either multiple starting points or multiple periods as repeatedly asserted in the Examiner's Answer. As explained by Appellants (Appeal Brief at p. 9), Bi uses its "look back" feature to calculate only one set of starting and ending points. This represents only one speech region, and Bi clearly does not describe identifying multiple speech periods – Bi doesn't need to and doesn't want to determine multiple periods for a single speech region. The Examiner's assertion to the contrary is without support in the reference and is simply not disclosed.

The repeated characterization in the Examiner's Answer of Bi's interim calculation points as "candidate starting points" for multiple periods also is misplaced. The interim calculation points are tested to see if the current SNR is less than the second, lower SNR threshold. Until that expression is found to be true, nothing is done at the interim calculation points; no speech period or starting point is generated. Only when the expression is found to be true, the algorithm sets "START", which "represents the actual starting point of the utterance." (Col. 6, line 15 to col. 7, line 30.)

#### **IV. THERE IS NO RATIONALE FOR COMBINING FUJII AND BI**

Interpreted correctly, Appellants' claims require "identifying a start position of a speech region of speech data for which speech recognition is to be performed." Interpreted correctly, a key feature of Fujii is not to identify a start position of a speech period, and Fujii does not disclose how specifically to determine the endpoints of its multiple proposed speech periods. Interpreted correctly, Bi discloses calculating a single set of start and endpoints for an utterance, does not calculate multiple start points, and does not identify multiple speech periods.

Thus, there is no reason to combine Fujii and Bi. Fujii's approach does not include identifying the endpoints of the speech period, yet the goal of Bi is to identify such endpoints. Moreover, Bi's endpoints are for only one speech period, yet a goal of Fujii is to generate multiple proposed speech periods.

When the references are interpreted correctly, the only remaining reason identified by the Examiner for combining them is simply that they "are analogous art because they are from a similar field of endeavor in speech recognition systems" (Examiner's Answer at p. 5). This argument is insufficient. As required by KSR (see Appeal Brief at pp. 8, 10) the Examiner must articulate some rationale for combining the references without resorting to hindsight analysis.

#### **V. THE EXAMINER'S ANALYSIS OF THE DEPENDENT CLAIMS IS INCORRECT**

Regarding dependent claim 3, the Examiner is incorrect in asserting that "there is no mention of the order of processing" in Appellants' claim as between identifying a start position of a speech region and generating a plurality of pieces of speech data (Examiner's Answer at p. 14, top). Particularly, the final limitation of claim 1 expressly recites that the plurality of pieces of speech data are generated by sequentially shifting back from "the start position of the speech region." Thus, the start position of the speech region is identified first, and the plurality of pieces of speech data are then generated on the basis of that start position. Again, this is contrary to the operation of Bi which identifies a single set of endpoints at the end of its "look back" process.

The Examiner's rejection of dependent claims 4, 12 and 20 is incorrect as for claim 3. In addition, the Examiner relies on the PRE\_START position in Bi as Appellants' "start position of the speech region," which is incorrect as explained above. Further, the Examiner cannot identify any disclosure in Bi that meets Appellants' claim limitation of "averaging speech data for several pieces of data from the start which have been subjected to the recognition processing." The Examiner points to the PRE\_START position in Bi, but that datum is only used once as an initial number in the

processing, is not "averaged" in any way, and does not correspond to the claimed "speech data for several pieces of data" in any event.

Regarding dependent claims 5, 13 and 17, the Examiner relies on his interpretation of Bi as generating "multiple candidate speech periods" and as "decrementing back from the start of a speech region." As explained above, both of these interpretations of Bi are flatly wrong, so that the rejection of these claims also is without basis.

## VI. CONCLUSION

The cited references do not disclose every limitation of Applicants' claims and cannot be combined in any reasonable fashion to result in Applicants' claimed invention. Therefore, the rejections should be reversed and the claims should be allowed.

Respectfully submitted,

6/19/09  
Date

James P. Naughton  
James P. Naughton  
Registration No. 30,665  
Attorney for Appellants

BRINKS HOFER GILSON & LIONE  
P.O. BOX 10395  
CHICAGO, ILLINOIS 60610  
(312) 321-4200